CHAPTER 5

Understanding and Contrasting the Data and Exploratory News Outlets

Felipe De La Hoz

Up until relatively recently, the idea of *data journalism* was a compelling yet somewhat mysterious concept in the news industry, a sort of luxury reserved for large newsrooms that could keep around a small stable of dedicated spreadsheet wranglers to perform little-understood data operations and report back their findings in plain English.

Though forms of computer-assisted reporting, or CAR, have been around since the 1960s, recognizable data reporting processes started becoming widespread and professionalized around the 1980s, as the technical tools available to reporters became more sophisticated and accessible. A quintessential example of the early successes in the field is the Pulitzer Prize-winning series on home lending discrimination by Bill Dedman of *The Atlanta Journal-Constitution*. The 1988 effort had Dedman, using techniques pioneered by researchers in academic departments—as well as collaborating with the researchers themselves—obtain data on home loans made over a six-year period in Atlanta and reported to the federal government. Based on this, he conclusively

F. De La Hoz (✉)
New York, NY, USA

O. Hahn and F. Stalph (eds.), *Digital Investigative Journalism*,
https://doi.org/10.1007/978-3-319-97283-1_5

determined that people in middle-income black neighbourhoods were less likely to get loans than counterparts in white neighbourhoods.

That series is very characteristic of the sort of data story of the time: the reporter got their hands on a set of data, in this case through filing a records request with the government; the reporter ran analysis on that data, extracting useful information and identifying patterns; and then the reporter wrote a straightforward news story using that data as one source, bolstering the conclusions drawn from his findings through speaking with other sources.

The picture of data journalism as a field quickly grew more convoluted. As technological advancements continued apace and recognition grew of the wide potential that computational tools unlocked, journalists and publishers began to realize that the classical methods and constraints of news presentation were increasingly less static, and to some degree irrelevant.

The proliferation of available and obtainable data was itself both an opportunity and a conundrum in a society where almost everything had started producing some data footprint, which now lay on top of the existing complexities of government, or business, or sports, or any other beat. Like data journalism, the concept of explanatory journalism had been around for decades, with its tenets first succinctly laid out by The Poynter Institute's Roy Peter Clark in a 1984 essay: write for a general audience, introduce concepts one at a time, use graphics and analogies.

Much of that new framework for crafting coverage sat unused until the sudden information onslaught of the modern information age forced its re-examination as a way to fulfil the core function of helping the society understand itself. It was newly dire to step away from daily, incremental coverage and attempt to present an issue in its relevant context distilled to a summation.

This, in turn, led to the question of the medium and format of the news; after all, the limitations had shifted from what could be printed on paper to what was technically achievable in an internet browser. It was a natural opportunity for data journalism specifically, which seemed uniquely positioned to pioneer new styles. While an interview can only really be presented as text (or videotaped), a series of related facts and figures, such as different sets of numbers per county for the Dedman stories, presents an immense array of possibilities when it comes to packaging. Do you have tabular data? Graphs? Interactive maps?

As part of this general removal of clutter, unnecessary expertise and analysis found itself on the chopping block. There was less need to try to

condense the opinions of an array of ostensibly knowledgeable sources into a meaningful narrative or assertion when the data itself could remove the uncertainty; no point having experts weigh in on whether or not historic patterns of discrimination were on obstacle to black families in Atlanta when you know which side is right.

The advent of modern data and explanatory journalism, then, was a combination of the impulses to condense complicated and information-heavy concepts simply and precisely, present that condensation in formats other than traditional articles—including previously unimaginable formats like interactive maps—and present it authoritatively, with the understanding that certain things are knowable and can be set forth as such.

Different notions of how to best apply these general principles has produced the variety in the outlets that we have come to associate with the genre. This chapter will focus on three of the largest and most fundamental: *The New York Times' The Upshot* vertical; *Vox*, a property of Vox Media; and *FiveThirtyEight*, an ESPN property.

To understand the distinctions between them, it is helpful to examine their inceptions. *FiveThirtyEight* is the oldest of these, having begun as the brainchild of statistician Nate Silver largely as a direct response to his personal dissatisfaction with campaign coverage in the run-up to the 2008 presidential election (Silver 2012). The foundational instinct, in this case, was a strong deference to the straightforward reliability of statistical methods in defiance of mountains of anecdotal, "soft" reporting.

Silver, who pioneered his data-driven style covering baseball, seemed to find no substantial difference between drawing conclusions from batting averages and earned runs over time and doing the same in political scenarios from historical polling and voter behaviour data. The unifying theme to the entire philosophy was that past is precedent, and smart insight is plainly waiting to be gleaned from diligent analysis of the information trail. Silver has not been shy about his disdain for pundits and punditry writ large, with its broad narratives concocted from the liberal interpretation of isolated incidents as opposed to more methodical extrapolation from available data.

Following the 2016 nomination of Donald Trump to be the Republican candidate for president, Silver even wrote a lengthy breakdown of what went wrong with his site's prediction (which had initially given Trump around a 5% chance) titled "How I Acted Like a Pundit And Screwed Up On Donald Trump" (2016). The gist of that piece is that Silver committed his own cardinal sin and presented

predictions without basis in the findings of a specific statistical model. *FiveThirtyEight* has thus remained, among its peers, the most "data-driven" of the outlets, with statistical interpretation continuing to make up its backbone.

It retains and is still largely known for its core concentration on politics (its name is derived from the number of electors in the US Electoral College). After it was started independently by Silver following a stint blogging for the popular Daily Kos in 2008, the site came under the auspices of *The New York Times*, where it was operated as a full-fledged vertical with the title "FiveThirtyEight: Nate Silver's Political Calculus." The site and the entire FiveThirtyEight brand was sold to ESPN in July of 2013 and has since expanded to include sections on politics, sports—a personal interest of Silver's, who began his blogging career at *Baseball Prospectus*—science and health, economics, and culture.

The site has always employed a general format of crisp, straightforward and information-heavy writing supplemented with simple graphics. It has generally eschewed flashy interactive graphical representations in favour of simpler, static visuals that are strict representations of data used to supplement the author's point. Certainly, writing styles differ between its sections, but the thrust throughout is informed analysis and, crucially, speculation.

That brings us to another of the key characteristics of the site: It is foundationally a predictive platform. Its most widely-known accomplishments at the start, after all, were calling elections. The site retains that spirit of forecasting, and many of its articles do feature commentary and prognostication while avoiding the kind of conventional wisdom of the operatives and gurus that Silver so disdains.

The Upshot actually began as a spiritual successor of sorts to *FiveThirtyEight* after Silver and his brand departed for ESPN. It was officially launched in April of 2014 under the direction of David Leonhardt, a long-time journalist and economics writer with the paper who, in a bit of foreshadowing, had launched the analytical sports column "Keeping Score." The top brass knew it couldn't let their blockbuster data vertical decamp to a competitor without replacement, but it was quickly decided that whatever came next could not just be a carbon copy of *FiveThirtyEight* (O'Donovan 2014). There had been grumblings of Silver's clashes with some of the Gray Lady's culture and established political reporting corps (Sullivan 2013); several people at the *Times* had felt that the purely analytical drive was incomplete and did not

incorporate enough "shoe-leather" reporting. *The Upshot* presented the leadership with an opportunity to take the lessons learned and rebuild their data vertical from scratch while trying to create a more equitable marriage between anecdote and data.

One of the most immediately noticeable things about *The Upshot* when compared to its predecessor is that it is attempting to communicate with a more diverse audience in a much more conversational tone. Indeed in his post introducing the site to readers, Leonhardt—who left the section for the opinion page in early 2016 and was replaced by Amanda Cox—wrote that the site would be written "in a direct, plain-spoken way, the same voice we might use when writing an email to a friend" (Leonhardt 2014). The site routinely addresses readers themselves, even presenting some of its articles in the frame of direct advice. One of the site's early hits was an interactive piece titled "Is it Better to Rent or Buy?" by Mike Bostock et al. (2014). The piece presented the reader with a series of interactive sliders that the reader would set to represent their own situation, and the site would automatically crunch the numbers and give the reader a rough recommendation on when they would be better off buying or renting a home.

In this way, *The Upshot* positions itself more firmly in the area of being explanatory and interpretative and is less focused on prediction than *FiveThirtyEight*. Though its pieces tend to incorporate data, not all are purely about data-crunching; the site features analyses of trends, explainers, and criticism, including dissections of other people's data work. Some read like straight news stories about topics that would be of interest to wonks and the technologically inclined, such as a recent datelined piece on how children are being prepared for an automated economy which featured the reporters going to a training program and interviewing parents and experts. This is likely a direct result of the vertical's desire to inhabit the space between pure data and pure shoe-leather.

The rent or buy piece illustrates another defining attribute of the site, which is an openness to experimenting with formats. Oftentimes, graphics are not supplementary but the journalism's main vehicle and they take many forms. Maps are a classic tool in frequent use, but they are joined by four-quadrant graphs, sliders, calendars, and assorted charts; many of them are interactive and responsive to the user. The rent or buy piece is also an example of a style of interactive by *The Upshot* where the reader is able to

enter data, such as by filling out a questionnaire or selecting from a series of options, and has the data crunched and a result spit back out to them.

If *FiveThirtyEight* is fundamentally data-driven and cerebral and natively predictive, *Vox* is the manifestation of a distinct, contrasting journalistic philosophy. It came into being in April 2014 as part of the larger Vox Media organization, which itself was a new company that began as the parent of the sports blog network SB Nation. It is the brainchild of journalist Ezra Klein, who was poached, along with several colleagues, from the *The Washington Post*'s *Wonkblog*, which Klein had founded as a policy and economics vertical within the *Post* with some of the same principles that he would then take to *Vox* (*Wonkblog* continues to operate at the *Post* and itself has its own particular characteristics and operating principles, though we will not delve into it here).

Klein had arrived at the *Post* from a background in blogging; he had developed a degree of wonkish policy expertise before joining the newspaper in 2009, and was given wide latitude to build *Wonkblog* up into a data-heavy traffic driver and must-read in D.C. political circles, with a roughly similar degree of autonomy as Silver had at *FiveThirtyEight*. He was lured away by Vox Media after failing to convince the leadership at the *Post* to provide him with a revamped platform that would have included an expanded staff (Byers and Gold 2014); after the move, Klein stated one of the catalysts was technology and the lack of a content management system that could allow robust storytelling in the way he was envisioning (Kaufman 2014).

It does not take a seasoned media analyst to identify the Vox's central intent. After all, it's right there in its tagline: Understand the News. It takes to heart the tenets of Clark's Poynter essay and essentially attempts a distillation of current events. *FiveThirtyEight* acts as a microscope of granular data crunching and prognostication, while *Vox* prefers to be a wide-angle lens trying to snapshot the whole picture of each consequential issue it tackles. It is no less an essentially data-focused enterprise but is not as preoccupied with the statistical dissection of each new development in an important story as it is with placing it in context. Perhaps the most concrete and celebrated manifestation of this are the site's *card stacks*, which are a kind of living interactive article that expands as stories develop; instead of a series of static articles or updates, the stack is more akin to a marriage between Wikipedia and a traditional piece of journalism. It will not just tell you that something happened and that it matters, but it aims to tell you why it matters, what historically set the stage for

it to matter, and how it might continue to matter. A notable example of this is the stack "Everything you need to know about Israel-Palestine" (Beauchamp 2017). The topic is certainly among the most multifaceted, impenetrable conflicts in modern history, with unending layers of causes and effects rooted in overlapping historical, religious, cultural, and martial contexts. A casual reader arriving at any isolated story on the Israeli–Palestinian conflict can expect to be hopelessly lost in the effort to grasp the development's significance for the larger state of affairs, and this in itself can be a powerful disincentive to dig deeper; why bother, it is practically Gordian.

So *Vox* truly starts at the beginning: the introductory card is "What are Israel and Palestine? Why are they fighting?" and the second card tackles "What is Zionism?" Even someone with no knowledge base on the subject can read and follow that. From there, each subsequent card builds on itself, arriving at topics like "Why are the US and Israel so friendly?" (card 13) and eventually getting to specifics such as "Why did Israel and Hamas go to war in July 2014?" (card 19). Ideally, a person could arrive with very little comprehension and leave having derived a decent understanding of the conflict's trajectory, both past and potentially where it could be headed in the future. By necessity, these overviews do not go very much in depth; each card tends to hover at around 400 words in length. This approach is not meant to supplant traditional news stories on ongoing developments on the topic, but rather to act as a constantly updating appetizer that will allow people to properly digest the news as it comes.

The site is also a breeding ground for experimentation in news presentation. Its explanatory essence leads it to use everything from standard graphs to cartoons in making news digestible for the unenlightened reader. Whatever the editors believe to be the most straightforward way to convey information is the medium they'll use. The card stacks are one example, but they are far from the most unconventional of *Vox*'s news presentations.

To get a more concrete sense of how the operational principles of each outlet shape their approach, it is useful to consider their handling of the same story. For our purposes, an ideal candidate is the 2017 debate over the repeal of the Affordable Care Act, specifically the Senate push to pass the so-called Graham–Cassidy bill to gut Obamacare. It was a prominent national story on a very complex topic involving lots of data and moving parts and intersected with health care, politics, taxes, and the

legislative heft of a federal administration newly controlled in its entirety by the Republican party. In short, the sort of thing that lends itself to plenty of deep, technical, and explanatory reporting.

The first *FiveThirtyEight* story to mention Graham–Cassidy was titled "Republicans Really Could Repeal Obamacare This Time," and written by senior political writer Perry Bacon, Jr. (2017). The piece is classic *FiveThirtyEight*, starting with the well-informed prognostication its title lays out. It is packaged as modules of information, such as section of brief bullet-pointed highlights of the bill's specific intended consequences; a quick explanation of why the bill is being introduced after previous efforts failed and why Republicans had to pass it before a September 30 deadline; and a detailed section on the likelihood of different Republican senators voting in favour, structured as four segments (almost certain yes, likely yes, almost certain no, likely no, and wild cards). It is number-heavy and detailed in a direct way, offering up analysis without wasting many words. For maximum effect, the reader would have to have some prior knowledge of the circumstances surrounding the legislative push; parachuting in would not quite work.

Another piece, "The New Health Care Bill Is Not So Great For States That Like Obamacare" by Anna Maria Barry-Jester (2017), was focused on taking a popular argument in favour the bill—that it would be uniformly beneficial for individual states in granting greater flexibility—and systematically blowing holes in it. It crunches the numbers on Medicaid funding changes and matches the information with states that expanded the Medicaid program under Obamacare, all presented using one of the site's characteristic multichromatic graphics. The piece finishes by pointing out that states with successful insurance exchanges will have to build up a new Medicare-replacement infrastructure with less money than they had before. In sum, it pairs data analysis with the institutional know-how to plainly assert what will happen, above the fray of politically motivated declarations of what *could* happen and inserts some predictions on the policy aftereffects of the bill's passing beyond the obvious.

True to form, coverage in the *Upshot* leans more towards the explanatory rather than a pure analytical system. Included in this is some of its hallmark pragmatist and utilitarian approach, such as discussing how the demise of the bill and the larger repeal effort could directly affect members of the public with a piece titled "How Failure of the Obamacare Repeal Affects Consumers," by Margot Sanger-Katz (2017b). The piece touches on but largely sets aside the political manoeuvring and the

big-picture policy implications and focuses on what the bill's death will mean in practice for different categories of people. For example, it reads "Those with employer insurance can also coast along with the status quo; there have been no policy changes affecting the price or comprehensiveness of workplace health plans."

As the legislative fight was ongoing, the vertical also attempted to meticulously dissect what would occur if Graham–Cassidy were to be passed, in "How the Latest Obamacare Repeal Plan Would Work," by Reed Abelson and Sanger-Katz (2017); in contrast to *FiveThirtyEight*'s method, it is more conversational and seems geared towards the layman. "The legislation also eliminates the tax penalty that people who refuse coverage face — the individual mandate — which could discourage insurance enrolment among healthier people, who are critical to making the program work," the authors explain. They go through some potential impacts of the block grant system, the spectre of a consumer protections rollback, and ramifications for Medicaid. It's designed to get the reader up to speed on what to expect if the bill becomes law, without diving too deep into each aspect.

The *Upshot* also cleverly exposed the political dynamics at play above and beyond the bill's technicalities by highlighting how many of the concerns raised by senators pushing the bill could have been addressed through provisions already in the ACA itself in a piece titled "One Way for G.O.P. to Achieve Some Repeal Goals? It's Already Part of Obamacare," by Sanger-Katz (2017a). In doing so, it further leads the reader to an understanding of the environment in which the debate is taking place, one in which not only outcomes matter but sometimes process and political points matter even more.

Vox took a slightly different approach; the site conducted saturation coverage of the Graham–Cassidy fight, with the sort of incrementalist approach that it was founded on. This includes at least three separate headlines containing the word "explained"—"Graham-Cassidy: The last GOP health plan left standing, explained," by Sarah Kliff, "The past 24 hours of Graham-Cassidy, explained," by Dylan Scott, and "Graham-Cassidy, the GOP's last health care proposal, explained with a cartoon," by Alvin Chang (2017). As its title suggests, that last one is formatted as a series of cartoons with accompanying explainer text. Chang uses the example of feeding birds as a representation of the different insurance systems, and keeps the cartoons and concurrent reporting simple and to

the point. It is meant for an audience that quite possibly had no idea how Obamacare or the Graham–Cassidy bill really worked.

The site also takes a separate tack in the tenor of the coverage, which often takes on a chattier, more first-person tone. "I've covered the GOP repeal plans since day one. Graham-Cassidy is the most radical" reads the headline of a Sarah Kliff piece (2017), which uses the author's impressions based on her own reporting trajectory combined with multiple data sources and experts to conceptualize for the reader what the likely impact of the bill's passage would be. Kliff also makes more analytical, subjective declarations not based strictly on numeric data, such as the assertion that the bill is moving along despite actually making issues in the previous bills worse, largely on the basis of it being a last-ditch effort. Another piece, "GOP Senators are rushing to pass Graham-Cassidy. We asked 9 to explain what it does" by Jeff Stein (2017), seems to have a bit of a tongue-in-cheek premise at the surface, but in reality, it provides the valuable frame of reference that legislators pushing the bill don't necessarily agree on or even know what it does. It is the sort of idea that seems too obvious to do, yet yields valuable background to the ongoing debate.

References

Abelson, R., & Sanger-Katz, M. (2017, September 20). How the Latest Obamacare Repeal Plan Would Work. *The New York Times*. Retrieved from http://www.nytimes.com.

Bacon, P., Jr. (2017, September 18). Republicans Really Could Repeal Obamacare This Time. *FiveThirtyEight*. Retrieved from http://www.fivethirtyeight.com.

Barry-Jester, A. M. (2017, September 20). The New Health Care Bill Is Not So Great For States That Like Obamacare. *FiveThirtyEight*. Retrieved from http://www.fivethirtyeight.com.

Beauchamp, Z. (updated 2017, May 1). Everything You Need To Know About Israel–Palestine. *Vox*. Retrieved from http://www.vox.com.

Bostock, M., Carter, S., & Tse, A. (2014). Is It Better to Rent or Buy? *The New York Times*. Retrieved from http://www.nytimes.com.

Byers, D., & Gold, H. (2014, January 21). Why the Post Passed on Ezra Klein. *Politico*. Retrieved from http://www.politico.com.

Chang, A. (2017, September 20). Graham-Cassidy, The GOP's Last Health Care Proposal, Explained with a Cartoon. *Vox*. Retrieved from http://www.vox.com.

Clark, R. P. (1984, January–February). Making Hard Facts Easy Reading: 14 Steps to Clarity. *Washington Journalism Review*. Retrieved from http://cdn.nycitynewsservice.com/blogs.dir/422/files/2014/02/MakingHardFactsEasyReading.pdf.

Dedman, B. (1988, May 1). Atlanta Blacks Losing in Home Loans Scramble; Banks Favor White Areas by 5-1 Margin. *The Atlanta Journal-Constitution*. Retrieved from http://www.powerreporting.com.

Kaufman, L. (2014, April 6). Vox Takes Melding of Journalism and Technology to a New Level. *The New York Times*. Retrieved from http://www.nytimes.com.

Kliff, S. (2017, September 20). I've Covered the GOP Repeal Plans Since Day One. Graham-Cassidy is the Most Radical. *Vox*. Retrieved from http://www.vox.com.

Leonhardt, D. (2014, April 22). Navigate News With The Upshot. *The New York Times*. Retrieved from http://www.nytimes.com.

O'Donovan, C. (2014, April 23). Q&A: David Leonhardt Says The Upshot Won't Replace Nate Silver at *The New York Times*. *Nieman Lab*. Retrieved from http://www.niemanlab.org.

Sanger-Katz, M. (2017, September 23). One Way for G.O.P. to Achieve Some Repeal Goals? It's Already Part of Obamacare. *The New York Times*. Retrieved from http://www.nytimes.com.

Sanger-Katz, M. (2017, September 26). How Failure of the Obamacare Repeal Affects Consumers. *The New York Times*. Retrieved from http://www.nytimes.com.

Silver, N. (2012, November 10). Why I Started FiveThirtyEight. *The Daily Beast*. Retrieved from http://www.thedailybeast.com.

Silver, N. (2016, May 18). How I Acted Like a Pundit and Screwed Up on Donald Trump. *FiveThirtyEight*. Retrieved from http://www.fivethirtyeight.com.

Stein, J. (2017, September 20). GOP Senators are Rushing To Pass Graham-Cassidy. We Asked 9 to Explain What It Does. *Vox*. Retrieved from http://www.vox.com.

Sullivan, M. (2013, July 22). Nate Silver Went Against the Grain for Some at The Times. *The New York Times*. Retrieved from http://www.nytimes.com.